REMARKS

Applicants and the undersigned are most grateful for the time and effort accorded

the instant application by the Examiner. The Office is respectfully requested to

reconsider the rejections presented in the outstanding Office Action in light of the

following remarks.

There were several objections to the specification in the outstanding Office

Action. Reconsideration and withdrawal of these objections is respectfully requested.

First of all, language appearing on Page 6, line 7 was objected to because a dash

was used instead of an equals sign. The paragraph containing the text in contention has

been amended to correct this issue. It is respectfully requested that this objection be

reconsidered and withdrawn.

Second, there were certain formatting issues with items 3 and 4 in Table 1. The

items were centrally aligned rather than aligned with the left border of the Table, and

further, there was no space between the item number and the item title. Items 3 and 4 in

Table 1 have been amended to address these formatting issues. Thus, the table can no

longer be interpreted as being in conflict with the statements set forth in the specification.

It is respectfully requested that this objection be reconsidered and withdrawn.

Finally, the specification is objected to because the term 'ANWE' allegedly lacks

antecedent definition or description. This rejection is respectfully traversed. Applicant is

simply using an acronym for his invention titled 'Automated New Word Extraction'.

- 8 -

Thus, 'ANWE' does not lack antecedent basis in the specification. Reconsideration and withdrawal of this objection is respectfully requested.

Claims 1-19 were pending in the instant application at the time of the outstanding Office Action. Of these claims, Claims 1, 10, and 19 are independent claims; the remaining claims are dependent claims.

Claims 1 and 10 stand rejected under 35 USC § 112, second paragraph as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Specifically, the outstanding Office Action asserts that the limitation "a cleaned corpus" lacks a clear scope in the claim because the specification allegedly does not describe or clearly define a level of type of 'cleanness' for a corpus. This rejection is respectfully traversed. It seems to be well-known in the art that "cleaning" a corpus includes suppressing the noise in a document (i.e., graphics and correlated text is suppressed). Numerous texts and scholarly documents recite and use cleaned corpuses or the process of cleaning corpuses in linguistics and language processing work. For example, a search at http://scholar.google.com on the key words "how to clean a corpus" yielded thousands of documents that utilized clean corpuses or explained corpus cleaning with regards to linguistics and language processing. Similar results were obtained using other search engines, especially those directed towards scholarly documents. Thus, reconsideration and withdrawal of this rejection is respectfully requested.

Claims 1-3, 6-12, and 15-19 stand rejected under 35 USC § 103(a) as being

unpatentable over Wang et al. (hereinafter "Wang") in view of Razin et al. (hereinafter

"Razin"). Specifically the Office asserted that "[i]t would have been obvious ... to

modify Wang by specifically providing filtering a set of extracted phrases and creating

(output) final phrases list, as taught by Razin, for the purpose of obtaining extracted

words constituting significant user phrases." Reconsideration and withdrawal of the

present rejections are hereby respectfully requested.

The present invention is directed to a method and apparatus for automatically

extracting new words from a cleaned corpus. The instant invention segments a cleaned

corpus to form a segmented corpus, splits the segmented corpus to form sub strings, and

counts the occurrences of each sub strings appearing in the given corpus. Finally, the

present invention filters out false candidates to output new words.

As best understood, Wang appears to be directed to a method that optimizes

language models in which an initial language model is developed from a lexicon and

segmentation derived from a received corpus. The initial model is iteratively refined by

updating the lexicon and re-segmenting the corpus using both maximum match

techniques and statistical principles. (Abstract) However, as asserted in the outstanding

Office Action, Wang does not expressly disclose filtering out false candidates to output

new words, nor does Wang simply output words. Rather, Wang produces and optimizes

language models.

Razin does not overcome the deficiencies of Wang as stated above. As best understood, Razin appears to be directed to standardizing phrasing in a document. Razin identifies phrases in a document to create a preliminary list of phrases, then filters and refines those phrases to create a final list of standard phrases. Razin then identifies phrase of a document that are similar to standard phrases, decides if the candidate phrase is similar enough to the standard phrase and compute phrase substitutions to determine the approximate conformation of the standard phrase to the approximate phrase and vice versa. (Abstract) Razin explicitly asserts that his invention uses a tree based on stemmed words and known elements, not character strings, as in the instant invention. (Col. 1, line 54 to Col. 2, line 2) Further, Razin outputs standard phrases, not new words in a document. Thus, there is no teaching in Razin to filter out false candidates in order to output new words.

Claim 1 recites, *inter alia*, filtering out false candidates **to output new words**. (emphasis added) Similar language also appears in the other Independent Claims. Neither Wang nor Razin, nor the combination of the two, teach or suggest the limitations of the instant invention.

Further, a 35 USC 103(a) rejection requires that the combined cited references provide both the motivation to combine the references and an expectation of success. Not only is there no motivation to combine the references, no expectation of success, but actually combining the references would not produce the claimed invention. Thus, the claimed invention is patentable over the combined references and the state of the art.

- 11 -

There is an inherent tension in Wang and Razin given that, in Wang, the invention deals with word-level trees and does not relate at all to the process of standardized document phrasing, which is the crux of the invention of Razin. In fact, Razin specifically references this tension in a document similar to Wang and teaches away from combining the two references. (column 2, lines 22-40) Additionally, Razin asserts that his invention uses a tree based on stemmed words and known elements, not character strings, as Wang. (Col. 1, line 54 to Col. 2, line 2) At best, however, combining Wang and Razin would result in producing a language model of phrases which includes a lexicon of standard phrases rather than words. Even if there were a motivation for the combination, this combination does not teach or suggest the claimed invention.

Claims 4-5 and 13-14 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et al. (hereinafter "Wang") in view of Razin et al. (hereinafter "Razin") and further in view of Hui. Specifically the Office asserted that "[i]t would have been obvious ... to modify Wang in view of Razin by specifically providing using extended suffix tree (GST or GAST), for the purpose of storing more than one input strings." Reconsideration and withdrawal of this rejection is hereby respectfully requested.

Hui does not overcome the deficiencies of Wang or Razin. As best understood, Hui is directed towards an algorithm that provides an optimal sequential solution of the color set size problem which entails finding the number of different leaf colors in a subtree rooted at a vertex v in a rooted tree. Although Hui asserts that there is applicability in string matching heuristics, there is no teaching or suggestion in Hui to filter false word candidates to output new words.
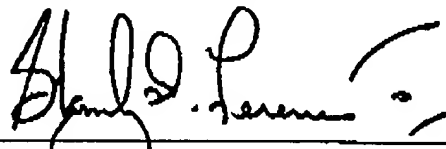
- 12 -

Combining Wang, Razin, and Hui would result in producing a language model of phrases using an optimal sequential solution to find the phrases that constitute the lexicon of standard phrases. Even if there were a motivation for the combination, this combination does not teach or suggest the claimed invention.

In view of the foregoing, it is respectfully submitted that Independent Claims 1, 10 and 19 fully distinguish over the applied art and are thus allowable. By virtue of dependence from Claims 1 and 10, it is thus also submitted that Claims 2-9 and 11-18 are also allowable at this juncture.

In summary, it is respectfully submitted that the instant application, including Claims 1-19, is presently in condition for allowance. Notice to the effect is hereby earnestly solicited. If there are any further issues in this application, the Examiner is invited to contact the undersigned at the telephone number listed below.

Respectfully submitted,

Stanley D. Ference III
Registration No. 33,879

Customer No. 35195
FERENCE & ASSOCIATES
409 Broad Street
Pittsburgh, Pennsylvania 15143
(412) 741-8400
(412) 741-9292 - Facsimile

Attorneys for Applicants

- 13 -